How to Bridge the Gap between AI Ethical Guidelines and Responsible Ethical Conduct

Uma G. Gupta
University of South Carolina Upstate
(ugupta@uscupstate.edu)

DOI: 10.26573/2023.17.1.4
Volume 17*, Number* 1
January 2023*, pp.* 41-53

Ethical violations, real or potential, within the ecosystem of Artificial Intelligence (AI) and its branches, have triggered global interest and concern about the unintended consequences of AI on humans. Although the number of AI ethical guidelines has exploded in recent years, there is a growing concern and preliminary evidence that guidelines alone are insufficient to protect individuals and society from harm. This conceptual research paper provides an overview of the importance of ethical AI guidelines as an entry point for building a foundation and framework within which human rights are protected, strengthened, and preserved. It outlines reasons why AI guidelines are often considered ethical washing and why guidelines are insufficient to hold organizations accountable. It identifies several reasons why AI is a "wicked problem" and why AI ethics may need a fresh design and comprehensive outlook and review. A few challenging and fundamental questions for future research are identified.

Keywords: Artificial Intelligence, AI Ethical Guidelines, Ethics Monitoring, Ethics Washing.

1. Introduction

"Artificial Intelligence refers to systems that can be designed to take cues from their environment and, based on those inputs, proceed to solve problems, assess risks, make predictions, and take actions (Morley et. al.)". AI ethics is a field of study seeking "a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies" (Leslie 2019). Our understanding of the economic power and societal implications of AI commercialization and its related repercussion on human rights, welfare, and well-being continue to evolve. Balancing a society's highest ethical standards and aspirations with the potential risk of serious, irreversible, and adverse impacts triggered by socio-technical systems on humanity has and will continue to be a challenge. This is because organizations and societies are limited in their abilities and capacity to plan, scope, manage, regulate, and predict human behavior. Governments and scientific communities inevitably find themselves on a perpetual learning curve as they try to balance "AI good" with "AI risks and damage". Therefore, meaningful, and rigorous AI ethical guidelines, laws, and

regulations are critical artifacts and essential elements of an ethical ecosystem if we are to balance technological advances with the welfare and well-being of humanity.

Ethics is a 'common currency' that binds communities and eco-systems of inventors, scientists, researchers, governments, educators, business leaders, policymakers, and practitioners in Artificial Intelligence (Cath 2018; Floridi and Strait 2020; Goirand et. al. 2021; Stix, 2021). Ethical violations have a global impact, with the consequences reverberating throughout the world in a matter of days or months. The COVID-19 pandemic serves as a harsh and compelling story of this unhappy truth. Today we are tightly coupled members of highly interlinked communities where the actions of one segment have a broad impact and deep consequences on entire societies, even distant ones. Our ethical principles and actions have a multiplier effect, both predictable and unpredictable. Therefore, shortcomings or weaknesses in AI ethical codes and guidelines can have severe, far-reaching, long-term, and both predictable and unpredictable global consequences on humanity and our environment. Hence, the study of ethics across scientific communities is no longer an academic endeavor, but a method and a mechanism for survival (Arnold 2018).

Today, there is a growing consensus that our global and interlinked ecosystem of AI products and services, innovations, discoveries, knowledge advancements, and creativity can only survive and thrive if societies and organizations are open and willing to be dedicated members of an interlinked community of learners and doers (Butcher and Beridze 2019; Chaturvedi 2015; Miller 2017, Watch 2019). In other words, the entire ecosystems of individuals, organizations, and entities involved in AI, whether it be at a cursory or deep level, must be open to transparency, influencing and being influenced, steered, nurtured, and monitored by universal and inviolable principles of ethics that transcend geography and cultures (Butcher et. al. 2019; Hagendorf 2020; Jobin 2019; Whittlestone et.al. 2019). Governments, policymakers, and watchdogs are therefore redirecting their focus on *evidence-based* commitment and proof that intelligent systems, products, and services uphold and protect the fundamental ideals and principles of human rights and welfare. This loud and growing call for compelling evidence or proof of adherence to strict and meaningful AI ethics is the result of a growing realization that simply hoping for the best outcome from AI products and services is a high-risk strategy with dangerous consequences. (Algorithm Watch 2020; Svetlova 2022). In this context, AI ethical guidelines are viewed as merely an entry point to an evidence-based ethical culture.

New advances in AI and its innovative applications and discoveries have created extraordinary openings and opportunities, both apparent and concealed, for ethical violations. Individuals, organizations, and even entire scientific communities tend to miss the challenges and moral violations that may lie hidden and inherent in scientific ideas, technological products, and innovative services, particularly those with promises of large commercial rewards. Commercial entities, such as technology companies, professional associations, government entities, and scientific communities, prompted by the desire for speed to market, may develop unsophisticated and linear rules for complex ethical decisions. (Morley et. al 2021; Morgan 2019). Such decisions have long-term and even severe consequences on segments of society or innocent bystanders. References. "AI ethics too often fails to account for the business practices—particularly, the business decision-making

practices—which form the normative basis for algorithmic decision-making." (Attard-Frost).

Hidden prejudices and biases can also derail leaders from actively and conscientiously exploring the potential risks and ethical violations of scientific innovations, discoveries, and products. Stakeholder demands are another key stress point in balancing ethics and profits *References*. The unrelenting compulsion, pressure, and urgency to open the floodgates to capitalism and thereby, quickly reap the economic benefits of inventions and discoveries are not just an issue of our times. It has been an issue that has dogged the scientific community and the business world for centuries and will continue to shape our future. (*References)* In short, **"**while principles are important, they are not in themselves enough to ensure society can reap the benefits and mitigate the risks of new technologies" Whittlestone.

2. What are Ethical Guidelines?

Ethical guidelines in scientific disciplines are a collation or codification of the collective thoughts, ideas, trends, experiences, and mental models of an organization or community, although such guidelines may vary across nations (Lee et. Baum, 2017; Zeng et.al., 2019; Winfield, 2018). Comprehensive guidelines that extend beyond addressing just technical issues, such as algorithmic bias, are designed to assist with the development, deployment, growth, and governance of intelligent applications that may have far-reaching consequences on individuals and societies (Cath 2018; Floridi and Strait 2019). Meaningful, yet practical, comprehensive, and rigorous ethical guidelines are fundamental to the growth and sustainability of AI applications, particularly when such systems proliferate into the everyday lives of ordinary citizens. Without such guidelines, it is difficult to ensure that the impact of AI products, services, and research are aligned with the social, economic, political, technological, and moral norms and well-being of a society (Algorithm Watch, 2020), particularly when intelligent systems present difficult situations and dilemmas (ACM U.S. Public Policy Council 2017, Weinbaum et.al. 2019).

In rapidly emerging technologies, particularly those that have a direct bearing on the everyday lives of citizens, guidelines are designed to be interpreted and scaled to the highest order of universal values. This includes values such as accountability, explainability, fairness, privacy, integrity, sustainability, environmental regard, transparency, nonexploitation, and concern for the welfare of the common good (Hagendorff, 2020) to name a few.

The Proliferation of AI Guidelines

The literature is rich with guiding principles, rules, regulations, tools, and compliance requirements on AI ethics in the form of research papers, mission statements, white papers, and other digital resources. (11, Lee et.al., Stix, 2020; Jobin et.al*;)* In 2018, the European Commission established the High-Level Expert Group on Artificial Intelligence to explore the space of AI ethical violations and establish meaningful guidelines. There have been several other comprehensive AI guidelines including The Montréal Declaration for Responsible Development of Artificial Intelligence. Ethics Guidelines for Trustworthy AI, Asilomar Artificial Intelligence Principles (Buruk, Banu – citation stored in Research folder). In recent years there

has been a proliferation of AI guidelines (Whittlestone et.al. 2019) with a significant and notable increase since 2016 (Jobin) although the evidence of adherence of such ideals still falls short (ACM U.S. Public Policy Council 2017; Fjeld, 2020; Jobin, 2019; Vold 2019). Jobin et. al., in their review of 84 AI ethical guidelines noted that private companies and government agencies were the largest generators of ethical guidelines with the United States leading other nations and the UK in second place.

Jobin et. al. conducted a review of 84 AI ethical guidelines and principles. In their paper, Fjeld et al. researched thirty-six leading AI ethical guidelines documents and identified eight value-based categories or "thematic trends" and forty-seven individual principles (15). Rand Corporation identifies ten ethical principles that bind scientists and their scientific work across disciplines and geography. It provides a framework to study and understand the ecosystem of ethics across scientific research and application domains (Floridi et. al. 2018).

The themes that emerge from these guidelines provide the foundation for intellectual and moral discussions about AI ethics. A quick review shows that there is a common ground between universal ethical themes and that which is unique to AI. For example, Hagendorff based on a study of twenty-two ethical guidelines notes that accountability, privacy, and fairness appeared in about 80% of all guidelines. Cowls et. al., classify AI ethical guidelines into five categories namely, autonomy, beneficence, non-maleficence, and justice (which are the same four ethical categories in biomedical ethics (Beauchamp), and the fifth category is explicability (Whittlestone et. al. 2019). In other words, AI companies, products, and services must at the very least, meet the *minimum* standards of ethical guidelines that are universally common to *all scientific communities.* These categories apply across all scientific communities and geographic locations. These common themes are a starting point to establish the first step on the ladder of elevating AI's professional conduct, ethics, and value to the highest order of universal values. An approach that focuses on the common ground can create focus and trigger a set of actions driven by social norms, cultural nuances, subtilities of geography and location, and the unique and emerging ethical challenges embedded within AI. It is understood that all principles require some level of customization, refinement, elaboration, and thoughtful deliberations by diverse stakeholders relevant to the product or services being offered. This is critical to building sophisticated cognitive, moral, social, economic, political, technical, legal, and cultural arguments of reasoning and nuanced thinking.

If there is an abundance of guidelines, why then is society concerned with the rapid advances of AI? Existing AI guidelines have several weaknesses that must be vigorously addressed (Stix 2020). Most principles are too general to be useful or practical when it comes to execution in a situation.

Guidelines are Insufficient

Ethical guidelines and frameworks are necessary for the advancement of a given discipline (Cowls 2019; Dameski 2018); however, they are insufficient as enforcement tools and mechanisms. Guidelines only provide a platform for discussion and debate, but frequently they do not translate into meaningful laws, regulations, policies, governance principles, or ethical conduct in the commercial world. "*The aggregate success of many AI Ethics Principles remains limited*

(Rességuier and Rodriques 2020).” There is a growing body of evidence that points to a chasm between well-intentioned guidelines and the ethical conduct of individuals, organizations, governments, and societies. (Morley et. al; Raji et. al. 2019; DePaoli 2014., Floridi 2019b.). Guidelines, therefore, are at best an entry point for ethical discussions, evaluations, and predictions of human behavior to resolve moral dilemmas. Guidelines do not prevent or dissuade ethical violations (Ville 2019. Conflicts and contradictions and discrepancies between principles are inevitable and likely. Whittlestone et.al. 2019 refers to these as “tensions” and identifies four tensions as it applies to AI principles and guidelines. 1) Data usage vs. privacy and autonomy b) making automatic decisions vs. fairness and equality c) benefits of increased individualization vs. solidarity and d) convenience vs. dignity and self-actualization. However, guidelines play an important role in that they provide a framework within which to anchor ethical decision-making. *“Without an agreed framework of norms that clearly identifies and articulates the relevant ethical standards which AI systems should be expected to comply with, little real progress will be made towards ensuring that these systems are in practice designed, developed, and deployed in ways that will meet widely accepted ethical standards. Yeung (2020)*

Hagendorff (2019) in his paper on the “The Ethics of AI Ethics: An Evaluation of Guidelines,” outlines the weakness and challenges of AI ethical guidelines and the importance of protecting society from harm, intentional or otherwise. . In an eye-opening experiment by McNamara et al. (20), they show that ethical guidelines neither inspire nor act as deterrents in influencing the ethical conduct of professionals in the tech community. In other words, the presence of ethical guidelines or the absence of it makes no difference in individual conduct. An “ideal hypothetical outcome is not the same as a workable pragmatic outcome” (Stix) “the current versions of AI Ethics Principles have fallen short of their promise and reached their limitation for impact in governmental policy-making” (Stix, 2020). Without enforcement mechanism, “AI research and development takes place in “closed-door industry settings”, where “user consent, privacy, and transparency are often overlooked in favor of frictionless functionality that supports profit-driven business models” writes Hagendorff (2019)

Self-monitoring and self-regulation are essential for corporations, governments, research labs, governments, and other entities, but in the absence of policing that is guided by far-reaching “laws with claws”, such mechanisms are insufficient to minimize, let alone ensure or guarantee that a community upholds the highest ideas (Morley et. al; Ville 2019). Ethics, in the context of AI, cannot be studied and monitored in backrooms or boardrooms. The guardians of AI ethics cannot work in isolation or select favorable situation-dependent rules or abstract ethical guidelines. A rear-view mirror assessment of the consequences of compromising ethics to technological advancement, intentionally or unintentionally, is insufficient. Hindsight only teaches us the lessons we have learned. Instead, we must embed the lessons of failure and ethical violations into laws, regulations, policies, and governance that regulate ethical human conduct, individuals and organizations are likely to treat ethics as feel-good marketing tactics. Hence, as new scientific and technological discoveries and advances in AI reshape our world, what is needed is not a proliferation of more ethical guidelines, but laws, regulations, and compliance

standards. (Redacted, 2018; Schiff et al., 2020)". Hence, it is only fair to make this a fundamental premise of all ethical guidelines: it is a natural and inevitable trait of human behavior to pursue economic gain over ethical advancement. Given this fundamental trait of human nature, how can scientific communities ensure that ethical guidelines moderate and control this behavior for the betterment of humanity? In other words, we must work within the realm of human nature and conduct, rather than assume that individuals will always seek to live up to the highest human ideals.

3. Ethics Washing in AI

Ethics washing is a term frequently used to refer to lip service to ethical conduct and behavior; it refers to an inauthentic and temporary set of tactics deployed to appease critical stakeholders and the public; it is "ethics theater" that uses language to promote and advocate self-regulation. The primary focus of ethics washing is to keep regulators and watchdogs at bay to avoid independent oversight that may result in penalties or stricter ethical controls. "Laws with claws" can lead to expensive lawsuits, reputational risk, market risk, financial risk, and the loss of critical talent and it is in the best interest of commercial entities to prevent such laws from going into effect with ethical washing (Comande 2018). It is a "mumble and fumble" approach that may become a platform to say "the right things but do the wrong things."

Ethics washing or "virtue signaling" (Middlestat) does not refer to unintended or unintentional consequences of one's actions, but intentional overlooking or avoidance of real or potential ethical violations in the pursuit of profits. The goal of ethics washing is to meet the lowest common denominator without running afoul of the law, to find innovative and creative ways to avoid admission to errors and suffer the consequences of ethical violations and to distract key stakeholders from focusing on evidence and hard facts about an organization's ethical conduct.

AI companies are often accused of ethics washing (References) given the evidence of prioritizing profits over ethics. This has led to a growing concern that organizations may hide behind ethical guidelines and overlook risk and potential harm to individuals and society and violations of human rights to pursue profits.

The critical question then is how do we prevent or minimize ethical washing? Recommendations include tightening the loopholes in the laws and regulations of AI products and services; strengthening the roles and responsibilities of watchdogs; establishing clear stage gates for product development and testing and giving the public a larger voice in this process. These are relevant, valid, and practical solutions; however, these approaches and tactics are insufficient. To expect that the measures in the above categories and related domains will solve the problem of AI ethical violations is naïve and simplistic. Perhaps it is time to revisit our philosophy, methodology, approach, and tactics to AI ethics for the following reasons.

A Naïve Approach to a Wicked Problem?

AI ethics is a wicked problem. What is a wicked problem? The concept and elements of a "wicked problem" was first enumerated by Horst Rittel and Melvin Webber, design scientists, to elaborate on the nature and scope of problems that defied clear logic, easy articulation, or risk-free solutions. Wicked problems have several characteristics, but what is relevant in this context is that wicked problems are

unique, their solutions cannot be experimented through trial and error since "each trial counts" according to the authors as "planners are liable for the consequences of the solutions they generate; the effects can matter a great deal to the people who are touched by those actions." Ethical development and execution of AI products and services may be classified as a wicked problem, and we outline the reasons below.

A Young and Emerging Field: Algorithmic decision-making is still a young and emerging field with dimensions and nuances that are difficult to fully comprehend and analyze. Intelligent systems are now embedded into our personal and professional workspaces, and the lines between human intelligence and machine intelligence are blurred. For example, many intelligent systems define and determine a fundamental human right, namely, access. AI systems and algorithms allocate who gets access (or is denied access) to what, when, and why. In other words, an algorithm has the power to determine human priorities and privileges. "The way a wicked problem is defined determines its possible solutions," according to Rittel and Webber. Hence, how we articulate priorities and privileges, for example, will define and shape the course of this emerging field and its ethical issues and challenges.

AI is too Big: The world of AI is complex, large, interconnected, interdisciplinary and continually evolving with fluid boundaries. AI applications and its sub-domains, such as machine learning, natural language processing, deep learning, robotics, neural networks, fuzzy logic, vision, speech recognition, are too big to be controlled with one set of rules or ethical guidelines. Even within a given sub-domain or space – say natural language – the range of applications and their associated datasets – create a plethora of rules. Prof Stuart Russell, the founder of the Center for Human-Compatible Artificial Intelligence at the University of California, Berkeley discusses the challenge and risk of our inabilities and myopic view to understand the inherent differences between these domains and how we apply them to solve real-world problems. In other words, we need to develop deep and meaningful ethical guidelines for each specialization or scientific discipline under the umbrella of Artificial Intelligence.

Lack of Consensus on AI Ethics: There is debate and lack of universal agreement about both what constitutes 'ethical AI', the nature and scope of ethical requirements, technical standards and best practices (Jobin); Frequently used terms in ethical guidelines lack universal definitions and agreement. But the very definition of these principles, their priorities and interpretations, and execution varies significantly (Jobin) not only between governments, but also between organizations, associations, and scientific communities. In their study of 84 documents on ethical guidelines, the following key ethical pillars emerged: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, dignity, sustainability, and solidarity. However, there is a lack of consistency and uniformity in our interpretation of these ethical foundations.

Unpredictable Technology: A technology must be stable and predictable for it to gain the trust of key stakeholders and users. *"The EU Commission stresses that "to gain trust, which is necessary for societies to accept and use AI, the technology*

should be predictable, responsible, verifiable, respect fundamental rights and follow ethical rules". AI technology is anything but predictable. This is because AI systems are a conglomeration of multi-layered embedded technical systems, which creates a "black box" environment where what happens inside the black box is not always consistent or predictable. This creates "massive zones of non-transparency remain, caused both by the sheer complexity of technological systems and by strategic organizational decisions." Hagendorff. Hence, individual case violations trigger rules and regulations that govern the future, rather than rules and laws that prevent these violations in the first place either due to oversight or due to the unpredictable nature of cutting-edge AI systems. As Hagendorff argues, "On the way from ethics to "micro ethics", a transformation from ethics to technology ethics, to machine ethics, to computer ethics, to information ethics, to data ethics has to take place."

From Principles to Practice: The will to translate AI principles into practice is inadequate resulting in a growing concern that AI principles are mostly attempts at ethics washing. Accountability and sanction mechanisms are often lacking and the ability to question or challenge the early stages of development of products and services that violate human rights are not available to the public or potential users. Balasubramanian et. al., present case studies on three companies and their approaches to ethics. Not only does the emphasis and focus on specific ethical values shift from product to product and company to company, but how they perceive and address the ethical challenges also differs.

We need a clear road map that maps each theme or principle, say fairness, for example, with the commercial development of the AI ecosystem. The focus must shift from the relentless generation of principles to exploring and examining how existing principles can be implemented within each sub-domain of AI. (Brent Middlestat; McMillan 2019). Principles must be based on situations, role-playing of what happens when individuals and organizations confront ethical situations along the different stages of product development, and the grey space that lies between the interplay of individuals, compromising ethical situations, and organizational priorities (Whittlestone et.al. 2019*).* In other words, the core elements of traceability, evidence, and proof of ethical conduct must be fully developed and nurtured (Yeung)

4. Conclusion

Ethics is not a static manifesto. It is a living organism deeply influenced and shaped by the best and worst of human behavior at different stages and phases in the evolution of a given technology and its applications. Ethical principles and guidelines are frequently drafted at the start of a scientific journey, discovery, or innovation; sometimes they emerge at the crossroads of an important discovery or its applications; at other times, it may be an afterthought when individuals and societies stare at the unthinkable, yet intentionally or inadvertently overlooked, consequences of the gaps and blind spots of existing ethical principles and guidelines. Regardless of where or how the journey of ethics begins, it is an arduous one because ethics is a dynamic, organic, interrelated, and intricate system of rules, principles, guidelines, regulations, standards, and laws that are shaped by individual and collective human behavior. Further, ethical principles and actions are influenced by geography,

culture, social pressures, societal values and the inevitable, indispensable, and enormous power of economics and profits.

The relevance of ethical guidelines is closely tied to feedback loops that are triggered and monitored by the behavior of individuals, emerging scientific discoveries and related products and services, social norms and values, and government policies, to name a few. In other words, ethics is a journey of continuous calibration and recalibration of our understanding of ethics, its applications to the real-world, and human behavior. It is driven by methods that clearly articulate to the scientific community how to translate policies and guidelines into practice and can lead to behavior modification and reflection when backed by laws, regulations and professional accountability (22).

The collective understanding and commitment of all scientists and scientific endeavors to the universal founding ethical principles is what binds the entire ecosystem of human creativity and technological progress. This is true regardless of geography, culture, methodology, or discipline. This is fundamental to the advancement of ethics within a given discipline, its intersection with other disciplines, and its ability to self-police and monitor emerging ethical metrics.

Many questions remain unanswered and will probably always have only incomplete and inadequate answers. What can we do to ensure that ethical guidelines inspire and moderate responsible ethical behavior? What steps can we take to create meaningful ethics for each sub-domain of AI? What aspect of this wicked problem should we tackle first to gain universal consensus? What is the role of culture and geography on appropriate ethics? How do we manage the sometimes-conflicting ethical guidelines that govern justice, privacy, transparency, beneficence, informed consent, accountability, safety, security, and surveillance, bias, fairness, nonexploitation, fairness, human rights, and non-discrimination?

5. References

1. ACM US Public Policy Council 2017. Statement on Algorithmic Transparency and Accountability. https://www.acm.org/binaries/content/assets/publicpolicy/2017_usacm_statement_algorithms.pdf
2. AI ethics and systemic risks in finance, AI and Ethics, 1-13 - January 2022, https://doi.org/10.1007/s43681-021-00129-1, Ekaterina Svetlova
3. Alglorithm Watch. (2020, April 30). AI Ethics Guidelines Global Inventory. Algorithm Watch. https:// inventory.algorithmwatch.org/
4. Arnold, T., & Scheutz, M. (2018). The “big red button” is too late: An alternative model for the ethical evaluation of AI systems. Ethics and Information Technology, 20(1), 59–69. https://doi.org/10.1007/ s10676-018-9447-7
5. Balasubramaniam, N., Kauppinen, M., and Kujala, S., “Ethical Guidelines for Solving Ethical Issues and Developing AI Systems,” In: Maurizio Morisio, Marco Torchiano, Andreas Jedlitschka (Eds.), Product-Focused Software Process Improvement - Lecture Notes in Computer Science, Vol. 12562, 331-346 - November 2020
6. https://doi.org/10.1007/978-3-030-64148-1_21

7. Baum, S. D. (2017). Social choice ethics in artificial intelligence. AI & SOCIETY. https://doi.org/10. 1007/s00146-017-0760-1
8. Butcher, J., & Beridze, I. (2019). What is the state of artificial intelligence governance globally? The RUSI Journal, 164(5–6), 88–96. https://doi.org/10.1080/03071847.2019.1694260
9. Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2133), 20180080. https://doi.org/10.1098/rsta.2018.0080
10. Clarke, R. (2019). Principles and business processes for responsible AI. Computer Law and Security Review. https://doi.org/10.1016/j.clsr.2019.04.007
11. Chaturvedi A. Ethical governance and society. SSRN Journal. 2015;
12. Cowls, J., and Floridi, Luciano. Prolegomena to a White Paper on an Ethical Framework for a Good AI Society. (2018).
13. Comandé, Giovanni, Unfolding the Legal Component of Trustworthy AI: A Must to Avoid Ethics Washing (September 10, 2020). Available at SSRN: https://ssrn.com/abstract=3690633 or http://dx.doi.org/10.2139/ssrn.3690633
14. Dameski A. A comprehensive ethical framework for AI entities: foundations. In: Iklé M, Franz A, Rzepka R, Goertzel B, editors. Artificial General Intelligence: 11th International Conference, AGI 2018, Prague, Czech Republic, August 22-25, 2018, Proceedings. Cham: Springer International Publishing; 2018. p. 42–51.
15. DePaoli L. Ethics in the field: contemporary challenges. JEA. 2014 Dec;17(1):49–53.
16. Fjeld, Jessica and Achten, Nele and Hilligoss, Hannah and Nagy, Adam and Srikumar, Madhulika, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI (January 15, 2020). Berkman Klein Center Research Publication No. 2020-1, Available at SSRN: https://ssrn.com/abstract=3518482 or http://dx.doi.org/10.2139/ssrn.3518482
17. Floridi L. Establishing the rules for building trustworthy AI. Nat Mach Intell. 2019 May 7;1(6).
18. Floridi, L. (2019b). Translating principles into practices of digital ethics: Five risks of being unethical. Philosophy & Technology. https://doi.org/10.1007/s13347-019-00354-x
19. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review. https://doi.org/10.1162/99608f92.8cd550d1
20. Floridi, L., & Strait, A. (2020). Ethical Foresight Analysis: What it is and Why it is Needed? Minds and Machines, 30(1), 77–97. https://doi.org/10.1007/s11023-020-09521-y
21. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., et al. (2018) AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations'. Minds and Machines 28(4), 689–707. https://doi.org/10.1007/ s11023-018-9482-5
22. Goirand, Magali, Austin, Elizabeth, Clay-Williams, Robyn, "Implementing Ethics in Healthcare AI-Based Applications: A Scoping Review, Science and Engineering Ethics, 27(5), 61 - September 2021 https://doi.org/10.1007/s11948-021-00336-3

23. Hagendorff T. The ethics of Ai ethics: An evaluation of guidelines. Minds and Machines. 2020;
24. Hoskinson S, Kuratko DF. The challenges of ethics and entrepreneurship in the global environment. Hoskinson S, Kuratko DF, editors. Emerald Group Publishing Limited; 2015.
25. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2
26. Lee, T. S. H., Liu, S. Y., Wei, Y. L., & Chang, L. Y. (2020, November). A Comparative Study on Ethics Guidelines for Artificial Intelligence Across Nations. In International Conference on Innovative Technologies and Learning (pp. 289-295). Springer, Cham.
27. Leslie, D.: Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. https:// www. Turing. ac. Uk / sites/ default/ files/ 2019-06/ understanding_ artificial_ intelligence_ ethics_ and_ safety. pdf (2019). Accessed 3 June 2020
28. McMillan, D., & Brown, B. (2019, November). Against ethical AI. In Proceedings of the Halfway to the Future Symposium 2019 (pp. 1-3).
29. Miller K. Can We Program Ethics into AI? [Reflections]. IEEE Technol Soc Mag. 2017 Jun;36(2):29–30.
30. Mittelstadt B. AI Ethics–Too principled to fail. arXiv preprint arXiv:190606668. 2019;
31. Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4
32. Morgan J. Ethical & Human-Centered AI. Figshare. 2019;
33. Morley J, Floridi L, Kinsey L, Elhalal A. A Typology of AI Ethics Tools, Methods and Research to Translate Principles into Practices.
34. Morley, Jessica; Ellhalal, A, Garcia, F; Kinsey, L., Mokdander, J., Floridi, L., Ethics as a Service: A Pragmatic Operationalisation of AI Ethics, Minds and Machines, 31(2), 239-256 - June 2021 https://doi.org/10.1007/s11023-021-09563-w
35. Orr, W., & Davis, J. L. (2020). Attributions of ethical responsibility by Artificial Intelligence practitioners. Information Communication and Society. https://doi.org/10.1080/1369118X.2020.17138 42
36. Oxborough, C., Cameron, E., Rao, A., Birchall, A., Townsend, A., & Westermann, C. (2019). Explainable AI: Driving Business Value through Greater Understanding. PWC. https://www. pwc.co.uk/audit-assurance/assets/explainable-ai.pdf
37. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing (pp. 33–44). https://doi.org/10.1145/3351095.3372873
38. Rességuier, A., & Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. Big Data & Society, 7(2), 85.

39. Rittel, H. W., & Webber, M. M. (1973). "Dilemmas in a General Theory of Planning." Policy sciences, 4(2), 155-169. https://www.cc.gatech.edu/fac/ellendo/rittel/rittel-dilemma.pdf.
40. Rolin K. Scientific community: A moral dimension. Social Epistemology. 2017 Sep 3;31(5):468–83.
41. Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (2020). What's next for AI ethics, policy, and governance? A global overview (pp. 153–158). https://doi.org/10.1145/3375627.3375804
42. Sherwin S. Developing a working knowledge of ethical principles in research. British Journal of School Nursing. 2015 Oct 2;10(8):402–3.
43. Stix, Charlotte, Actionable Principles for Artificial Intelligence Policy: Three Pathways, Science and Engineering Ethics, 27(1), 15 - February 2021, https://doi.org/10.1007/s11948-020-00277-3
44. Tatebe J. The ethics of difference: ethical dilemmas of external researchers. J Acad Ethics. 2015 Sep;13(3):229–37.
45. Thomsen, K. (2019). Ethics for artificial intelligence, ethics for all. Paladyn, 10(1), 359–363. https:// doi.org/10.1515/pjbr-2019-0029
46. Torresen J. A review of future and ethical perspectives of robotics and AI. Front Robot AI. 2018 Jan 15;4.
47. University of Montreal 2017. Montreal Declaration on Responsible AI. https: //www.montrealdeclaration-responsibleai.com/the-declaration
48. Venturebeat 2019 https://venturebeat.com/2019/07/17/how-ai-companies-can-avoid-ethics-washing/
49. Ville, V., Kai-Kristian, K., & Pekka, A. (2019). AI ethics in industry: A research framework. 2505, 49–60
50. Vold K.V, Hernandez-Orallo J. AI Extenders: The Ethical and Societal Implications of Humans Cognitively Extended by AI. Apollo - University of Cambridge Repository. 2019;
51. Watch A. AI Ethics Guidelines Global Inventory. 2019;
52. Weinbaum C, Landree E, Blumenthal M, Piquado T, Gutierrez C. Ethics in scientific research: an examination of ethical principles and emerging topics. RAND Corporation; 2019.
53. Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19). Association for Computing Machinery, New York, NY, USA, 195–200. DOI:https://doi.org/10.1145/3306618.3314289
54. Winfield, A. F. T., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2133), 20180085. https://doi.org/10.1098/rsta.2018.0085
55. Yeung, K. (2020). Recommendation of the Council on Artificial Intelligence (OECD). International Legal Materials, 59(1), 27-34. doi;10.1017/ilm.2020.5
56. Zeng, Y., Lu, E., Huangfu, C.: Linking artificial intelligence principles. In: Espinoza, H., Héigeartaigh, S.Ó., Huang, X., Hernández-Orallo, J., Castillo-Effen, M. (eds.) Proceedings of the AAAI Workshop on Artificial Intelligence Safety, vol. 2301, pp. 103–106. Creative Commons CC0, Honolulu (2019)

About Our Author

Uma G. Gupta is the Director of Business Analytics at the University of South Carolina Upstate. She has published three textbooks in Information Systems, edited a book on expert systems, and co-authored a book on Data Governance. She has authored more than ninety research papers and conference proceedings. Her areas of expertise include Enterprise-wide Business Analytics, Data Literacy, Data Visualization, and Organizational Digitization. She is a certified Tableau Desktop Specialist. She holds a Ph.D. in Industrial Engineering and an MBA from the University of Central Florida and a Master's in Mathematics from Stella Maris College, India.